Geographic information retrieval

Purves, Ross S ; Jones, Christopher B

Geographic Information Retrieval

Ross Purves[1] and Christopher Jones[2]
[1]Department of Geography, University of Zurich
[2]School of Computer Science & Informatics, Cardiff University
ross.purves@geo.uzh.ch| c.b.jones@cs.cf.ac.uk

Introduction

This special issue of SIGSPATIAL Special presents a series of notes describing the state of the art in Geographic Information Retrieval. The notes are intended to provide a review of some of the challenges presented as key research areas in Geographic Information Retrieval [Larson, Jones and Purves], and reflect progress in the field in the intervening years. The challenges as originally set out in [2] were the following:

- detecting geographical references in the form of place names and associated spatial natural language qualifiers within text documents and in users' queries;
- disambiguating place names to determine which particular instance of a name is intended;
- geometric interpretation of the meaning of vague place names, such as the 'Midlands' and of vague spatial language such as 'near';
- indexing documents with respect to their geographic context as well as their non-spatial thematic content;
- ranking the relevance of documents with respect to geography as well as theme;
- developing effective user interfaces that help users to find what they want; and
- developing methods to evaluate the success of GIR.

These challenges reflect a typical process chain for carrying out GIR, using a wide range of methods from different domains, and all except user interfaces are represented in this issue. Furthermore, the increasingly prominent domain of mobile search and the influence of social networks on GIR are also considered. The delivery of services relevant to a mobile user's position and context has brought with it a range of new opportunities for GIR which differ from those provided by traditional desktop computers, for example in considering the user's movement through an environment in focussing a query or the need for delivering very focussed results to a small area of screen real estate. The rise of social networks, and the related possibilities for their exploitation in, for example, exploring the use of vague place names or understanding how users describe place in informal text, brings with it a whole new set of challenges for the application of methods from GIR.

In bringing together this special issue, we asked for contributions from a wide range of those active in the field. Many of the notes are authored by those who have recently completed PhDs in the field, with others being contributed by authors who have been active in the research area since its inception. Each note aims to bring a different perspective on a research area, and as such should present a distinct view of progress in the field and the state of the art. Two researchers were initially invited to contribute a note in each area and given the choice of either collaboratively authoring a note (e.g. Leidner and Liebermann), or preparing two notes providing complementary, and sometimes differing, viewpoints.

We hope that the complete set of notes will provide a valuable source for those starting out in the field, seeking an update on recent progress, or looking for new research challenges and look forward to seeing further progress in GIR in coming years.

Notes in this special issue

Leidner and Liebermann combine to explore issues in the detection of geographical references and, crucially, associated text. Thus, they explore a range of potential approaches to the geoparsing of not only explicit toponyms, but what they term complex geographic phrases such as "30km west of Zurich". One key issue identified by Leidner and Liebermann is ambiguity.

Ambiguity forms the focus of the notes from both **Overell** and **Buscaldi**. Overell firstly sets out what forms ambiguity typically takes, before considering why it is an important problem in practical applications and discussing both the significance of the problem and the types of information which can be used to address it. Buscaldi's note then discusses in more detail approaches to toponym disambiguation, and crucially the evaluation of toponym disambiguation methods. The lack of a standardised data set for testing of disambiguation methods (and thus also approaches to geoparsing in general) is considered by Buscaldi to be an important issue, and he lists the resources which are currently available for such a task. One issue discussed in the notes from Leidner and Liebermann, Overell and Buscaldi, is the issue of vernacular and vague place names.

Such toponyms are the subject of the note from **Schlieder and Henrich**, who contribute a nuanced discussion to the issue. Crucially they point out how vagueness in regions is not *per se* a property of a particular region, but rather relates to the nature of the task being performed by a user. **Schockaert's** note sets out a range of approaches to the problem of effectively modelling vague regions using, for example, methods related to both possibility and probability theory.

In the next note **Leveling** addresses the problem of indexing geographical information which has been grounded, with particular reference to the need for recognition of the multiple forms of geographical references in text, the potential of integration with semantic indexing and question answering techniques and the intelligent interpretation of spatial terminology in user queries.

In the context of relevance ranking, **Cai** reviews the use of geometric and thematic similarity measures and probabilistic methods for combining multiple measures. He notes the need for more evaluation of the effectiveness and efficiency of alternative measures, as well as the so far relatively unexploited role of context in geographical information retrieval.

Larson echoes some of these issues with a focus on probabilistic methods employing logistic regression and he elaborates on the details of some of the methods for determining similarity between geometric footprints of documents and for combining spatial and textual similarity measures.

Mandl's note provides a very useful overview of the multilingual evaluation exercises undertaken in GeoCLEF. This activity formed a key part of the learning experience of many of those active in GIR, and the resources developed are important in understanding why in general GIR systems did not exhibit better performance on average than standard *ad hoc* retrieval approaches. **Cardoso's** note attempts to explain these differences in more detail, suggesting that the general hypothesis that GIR will always perform better given a geographic query is likely to be very closely linked to, for example, the nature of the corpus on which retrieval is being performed and the qualities of the query itself.

Naaman highlights the immense potential of Web 2.0 social media to provide a rich source of spatio-temporal information about the nature of regions (districts) and landmarks, the paths that

people follow and the activities in which they engage. In a complementary article, **Murdock** challenges us to realise the strong social and geographical bias of such data, reflecting as it does the lives of a relatively wealthy "technorati" living and travelling predominantly in North America and Europe.

The first note concerned with mobile GIR, by **Reichenbacher and De Sabbata**, contrasts geographical relevance with conventional IR notions of document relevance, drawing attention to the user's engagement with physical entities in the environment and the importance of the user's mobility and spatio-temporal context.

In the final note in this special issue, **Mountain** elaborates on these distinctive issues with emphasis upon the concept of context and its multiple facets with regard to location and patterns of behaviour, the characteristics of the physical environment and its affordances, the constraints of mobile devices and the user's activities, interests and preferences.

References

[1] Larson, R. 1996. Geographic Information Retrieval and Spatial Browsing. In *GIS and Libraries Patrons, Maps and Spatial Information*, Smith, L. and Gluck, M. (eds.), Urbana-Champaign: University of Illinois, 81-124.

[2] Jones, C. B. and Purves, R. S. 2008. *Geographical Information Retrieval*. International Journal of Geographical Information Science, 22(3), 219-228.